Genetic therapy and genetic modification using neocentromeric minichromosomes

BACKGROUND OF THE INVENTION

FIELD OF THE INVENTION

The present invention relates generally to the field of genetic and proteomic therapy in mammals, avian species, plants and other higher organisms. More particularly, the present invention provides a target region within a mammalian, avian, plant or other eukaryotic chromosome or an artificial or engineered chromosomal construct which is capable of carrying and expressing a heterologous gene or other genetic molecule of interest. Even more particularly, the gene or genetic molecule of interest is expressed in a region of the chromosome which corresponds to or which immediately adjoins or is proximal to a centromeric or neocentromeric region or a functional derivative thereof or a latent, synthetic or hybrid form thereof. The target region may be in a cell's chromosome or in an artificial or engineered chromosome or mini-chromosome. The present invention further contemplates a method for facilitating genetic therapy or genetic modification or other applications including protein production for proteomic therapy in a mammal, avian species or plant or other higher eukaryotes, by introducing DNA into a centromeric or neocentromeric region or a region immediately joining or proximal thereto within a mammalian, avian species or plant or other higher eukaryote chromosome or an artificial or engineered chromosomal construct.

DESCRIPTION OF THE PRIOR ART

Bibliographic details of the publications referred to by author in this specification are collected at the end of the description.

Reference to any prior art in this specification is not, and should not be taken as, an acknowledgment or any form of suggestion that this prior art forms part of the common general knowledge in any country.

The rapidly increasing sophistication of recombinant DNA technology is greatly facilitating research and development in the medical and allied health fields. A particularly important area is in mammalian including human genetics and the molecular mechanisms behind some genetic abnormalities. Progress in research in this area has been hampered by a lack of a full understanding of the transcriptional potential of a centromere and the limited availability of cloned nucleic acid molecules encompassing a human centromere.

The centromere is an essential structure for sister chromatid cohesion and proper chromosomal segregation during mitotic and meiotic cell divisions. The centromere of the budding yeast *Saccharomyces cerevisiae* has been extensively studied and shown to be contained within a relatively short DNA segment of 125 bp that is organized into an 8 bp (CDEI) and 26 bp (CDEIII) domain, separated by a 78 to 87 bp, highly AT-rich, middle (CDEII) domain (Clarke and Carbon, *Annu. Rev. Genet. 19:* 29-56, 1995). The centromere of the fission yeast *Schizosaccharomyces pombe* is considerably larger, ranging from 40 to 100 kb and consists of a central core DNA element of 4 to 7 kb flanked on both sides by inverted repeat units (Steiner *et al., Mol. Cell. Biol. 13:* 4578-4587, 1993). The functional DNA components of a higher eukaryotic centromere have been characterized in a mini-chromosome from *Drosophila melanogaster* and shown to consist of a 220 kb essential core DNA flanked by 200 kb of highly repeated sequences on one side (Murphy and Karpen, *Cell 82:* 599-609, 1995).

The mammalian centromere, like the centromeres of all higher eukaryotes studied to date, contains a great abundance of highly repetitive, heterochromatic DNA. For example, a typical human centromere contains 2 to 4 Mb of the 171 bp α-satellite repeat (Wevrick and Willard, *Proc. Natl. Acad. Sci. USA 86:* 9394-9398, 1989; Wevrick and Willard, *Nucl. Acids. Res. 19:* 2295-2301, 1991; Trowell *et al., Hum. Mol. Genet. 2:* 1639-1649, 1993), plus a smaller and more variable quantity of a 5 bp satellite III DNA (Grady *et al., Proc. Natl. Acad. Sci. USA 89:* 1695-1699, 1992; Trowell *et al.* [1993] *supra*). The role of these satellite sequences is presently unclear. Transfection of a cloned 17 kb uninterrupted α-satellite array into cultured simian cells (Haaf *et al., Cell 70:* 681-696, 1992) or a 120 kb

α-satellite-containing YAC into humand hamster cells (Larin *et al., Hum. Mol. Genet. 3:* 689-695, 1994) appear to confer centromere function at the sites of integration. Other workers have analyzed rearranged Y chromosomes (Tyler-Smith *et al., Nature Genet 5:* 368-375, 1993), or dissected the centromere of the human Y chromosome with cloned telomeric DNA (Brown *et al., Hum. Mol. Genet. 3:* 1227-1237, 1994) and suggested that 150 to 200 kb of α-satellite DNA plus several hundred kb of adjacent sequences are associated with human centromere function. In addition, a human X-derived mini-chromosome that retained 2.5 Mb of α-satellite array has been produced by telomere-associated chromosome fragmentation (Farr *et al., EMBO Journal 14:* 5444-5454, 1995). In all these studies, it is not known whether non-α-satellite DNA sequences are embedded within the centromeric site and operate independently of, or in concert with, the α-satellite DNA.

In mammals, four constitutive centromere-binding proteins, CENP-A, CENP-B, CENP-C and CENP-H, have been characterized to varying extents and implicated to have possible direct roles in centromere function. CENP-A, a protein localized to the outer kinetochore domain, is a centromere-specific core histone that shows sequence homology to the histone H3 protein and serves to differentiate the centromere from the rest of the chromosome at the most fundamental level of chromatin structure - the nucleosome (Choo *Dev. Cell 1:* 165-177, 2001). CENP-B, a protein which associates with the centromeric heterochromatin through its binding to the CENP-B box motif found in primate α-satellite and mouse minor satellite DNA, probably has a role in packaging centromeric heterochromatic DNA - a role which, however, is not indispensable since the protein is undetectable on the Y chromosome (Pluta *et al., Trends Biochem. 15:* 181-185, 1990), is found on the inactive centromeres of dicentric chromosomes (Earnshaw *et al., Chromosoma 98:* 1-12, 1989), and whose gene can be knocked out in mice without detectable consequences to mitotic and meiotic cell divisions (Hudson *et al., J. Cell Biol. 141:* 309-319, 1998). CENP-C has been shown to be located at the inner kinetochore plate and has an essential although yet undetermined centromere function as seen, for example, from inhibition of mitotic progression following microinjection of anti-CENP-C antibodies into cells (Bernat *et al., J. Cell. Biol. 111:* 1519-1533, 1990; Tomkiel *et al., J. Cell. Biol. 125:* 531-545, 1994), from

its association with the active but not the inactive centromeres of dicentric chromosomes (Earnshaw *et al.* [1989] *supra*; Page *et al., Hum. Mol. Genet. 4:* 289-294, 1995; Sullivan and Schwartz, *Hum. Mol. Genet. 4:* 2189-2197, 1995), and from an embryonic lethal phenotype in *Cenpc* gene knockout mice (Kalitsis et al., *Proc. Natl. Acad. Sci. USA 95:* 1136-1141, 1998). CENP-H is the latest essential constitutively binding centromere protein that has been described (Sugata et al., *J. Biol. Chem. 274:* 27343-27346, 1999; Sugata *et al., Hum. Mol. Genet. 9:* 2919-2926). More recently, a new role for the mammalian centromere as a "marshalling station" for a host of "passenger proteins" (such as INCENPs, MCAK, CENP-E, CENP-F, 3F3/2 antigens and cytoplasmic dynein), has been recognized (review by Earnshaw and Mackay, *FASEB J. 8:* 947-956, 1994 and Pluta *et al., Science 270:* 1591-1594, 1995). These passenger proteins, whose appearance at the centromere is transient and tightly regulated by the cell cycle, provide vital functions that include motor movement of chromosomes, modulation of spindle dynamics, nuclear organizations, intercellular bridge structure and function, sister chromatid cohesion and release and cytokinesis.

U.S. Patent No. 6,265,211 and International Patent Publication No. WO 98/51790 describe an unusual human marker chromosome, mardel(10), which is 100% stable in mitotic division both in the human subject from which it was isolated and in established fibroblast and transformed lymphoblast cultures. A region of the mardel(10) chromosome has been cloned together with the corresponding region from a normal human subject. The nucleic acid molecules cloned contained no α-satellite repeats yet are mitotically stable. The nucleic acid molecules encompassed, therefore, a new form of centromere referred to as a "neocentromere". The centromeric regions of higher organisms have traditionally been described as inhibitory to transcriptional activity (Choo *Dev. Cell. 1*: 165-177, 2001). The large tracts of repetitive DNA found at centromeres have, until the advent of the present invention, prevented proper analysis of transcriptional activity. In accordance with the present invention, the arrangement of centromeric chromatin domains has been determined and expression analysis has identified transcription activity. This provides, therefore, a target for inserting genes and other genetic material into the chromosome of eukaryotic cells as well as in artificial or engineered chromosomal constructs.

SUMMARY OF THE INVENTION

Throughout this specification, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element or integer or group of elements or integers but not the exclusion of any other element or integer or group of elements or integers.

The present invention is predicated in part on an analysis of whether centromere formation is inhibitory to transcriptional activity in a higher organism or plant. This analysis is necessary in the design of genetic therapeutic or improvement protocols including the use of artificial or engineered chromosomes or in chromosomal genetic targeting or their use to produce proteins for proteomic therapy. A human neocentromere is used in accordance with the present invention which is amenable to functional dissection due to:

(a) the lack of large arrays of tandemly repeated DNA;

(b) the availability of fully sequenced markers which allows high-resolution molecular mapping of chromatin domains; and

(c) the presence of many naturally occurring genes on the same chromosomal region pre- and post-neocentromere formation.

Figure 1e summarizes the distribution of the different domains defined at the 10q25 neocentromere in mardel(10). The non-overlapping location of the CENP-A and HP1 domains is similar to what occurs in yeast and fly, where the core CENP-A-binding DNA has been shown to be devoid of heterochromatin proteins (Choo, *Dev. Cell. 1*: 165-177, 2001). The CENP-A and HP1 domains form the proximal and distal boundaries of a 1 Mb-region of delayed replication previously mapped at the 10q25 neocentromere (Lo *et al., EMBO J. 20:* 2087-2096, 2001A) (see Figure 1e), implicating a role of these domains in blocking the spreading of the acquired region of delayed replication timing.

A region of neocentromere-induced chromatin reorganization has been defined at 10q25 totalling approximately 4 Mb in length (Figure 1e). Aside from the uncertainty of gene silencing effects at the 330-kb CENP-A- and 100-kb and HP1-associated domains, the data clearly demonstrate that normal transcription is permissible for many different genes distributed over the remaining scaffold-enriched region, as well as within the 900 kb CENP-H domain and the 1 Mb region of delayed replication (Figure 1e). These data include the substantial remodification of chromatin that bestows the critical structural and functional properties of an active kinetochore has not in any measurable way compromised the transcriptional competency of the greater part of the underlying centromeric DNA, or accessibility of the cell transcriptional machinery to this DNA.

The present invention provides, therefore, a target for gene or other genetic material insertion either in a cell's chromosome or in an artificial or engineered chromosome. This target comprises a centromeric or neocentromeric region or immediately adjoining regions including proximal genetic locations.

The present invention permits the development of genetic therapeutic or genetic improvement protocols for mammals as well as proteomic therapeutic protocols including humans, avian species, plants and other higher organisms. In particular, artificial or engineered chromosomes carrying genetic material is operably linked to a promoter inserted within a centromeric or neocentromeric region or in an adjoining or proximal region. The genetic material is expressible when introduced into a cell and, hence, the artificial or engineered chromosome can be used to introduce or modify a phenotypic trait of a cell or an organism or plant carrying the cell, or to produce therapeutically useful proteins for large-scale extraction or for use in proteomic therapy.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 is a diagrammatic representation of the organization of the neocentromeric domain. **(a)** BAC array spanning 8 Mb showing position of clones used in CIA and SIA analyses. **(b)** Scaffold/matrix attachment along 10q25 BAC array as determined by SIA analysis on chromatin prepared from 5f (mardel(10) chromosome) and 1f (normal chromosome 10) hybrid cell lines. **(c)** Distribution of CENP-H antigen along 10q25 BAC contig (x-axis) as determined by CIA analysis. **(d)** Distribution of HP1α antigen along the 10q25 BAC contig (x-axis) as determined by CIA analysis. **(e)** Summary of domain-distribution properties at the 10q25 neocentromere.

Figure 2 is a representation showing CT analysis of expressed genes in the region of the 10q25 neocentromeric activation. Refer also to Table 1 for explanation of CT and ΔCT. 1/ΔCT (y-axis) provides a measure of expression level of individual genes. Comparison of results for somatic cell hybrids containing human chromosome 10 (grey bars) and those for hybrids containing mardel(10) (black bars) indicate no major difference between hybrid pairs for genes tested. Refer to Table 1 for a summary of the relative expression levels between hybrids and student's *t*-test values.

Figure 3 is a visualized representation of scaffold attachment on metaphase chromosomes. **(a, d)** FISH using BAC clones BA313D6 and BA427L15 which mapped outside the S/MAR-enriched domain identified by SIA analysis, produces dispersed signals (open arrows) on both the normal chromosome 10 (top panel) and mardel(10) chromosomes (bottom panel), indicating predominantly non-scaffold attachment of the probed regions. **(b, c)**. FISH using BAC clones E8 and BA153G5 mapping within the S/MAR domain produces dispersed signals (open arrow) on chromosome 10 (top panel) but tightly packed signal on the mardel(10) chromosome (closed arrow; bottom panel), indicating predominantly scaffold attachment of the probed regions on mardel(10).

Figure 4 is a representation showing truncation of mardel(10) in mouse embryonic stem cells. **(a)** Structure of TACT (telomere-associated chromosomal truncation) targeting

constructs used for truncating mardel(10). Targeting DNA (B43all and B79e16) from the p and q arm of mardel(10) [see (4b)] and a mammalian selectable marker (either puromycin or hygromycin resistance gene, puromycin [registered trademark] hygromycin [registered trademark]) were cloned into vectors containing small arrays of cloned human telomeric DNA (Htel). Constructs were linearized at a restriction site between the vector DNA and the telomere repeats to expose the telomere sequences at the terminal. **(b)** Schematic formation of mardel(10) and NC-MiCs derived from truncation of mardel(10). Open arrows indicate the breakpoints on the normal chromosome 10 in the generation of mardel(10). The long and short arms of mardel(10) are denoted as q and p, respectively. NC-MiC53g was formed as a result of truncation and deletion both the and q arms of mardel(10). NC-MiC8a and 20f were the result of truncations using construct targeting B43all site with puromycin resistance gene followed by a second truncation using construct containing hygromycin resistance gene targeting B79e16s. Vertical shaded area represents the centromere protein CENP-A-binding domain (Lo *et al.* [2001A] *supra*). Open arrowheads indicate positions of intended targeted truncation. (+) denotes a positive FISH result for a BAC or cosmid probe on an NC-MiC, while (–) indicates a negative FISH result.

Figure 5 is a photomicrographic representation showing FISH analysis of NC-MiC6 in human HCT116pgrxr cell line. NC-MiC6 is indicated by arrow and human chromosome 10 by arrowhead. **(a)** FISH using **(i)** neocentromeric probe B153g5 and **(ii)** p-arm probe B326h7 (green), showing the transfer of NC-MiC6 into HCT116pgrxr. **(b)** Split images of (a) showing DAPI staining.

Figure 6 is a photomicrographic representation showing FISH analysis of NC-MiC6 in human 293trex cell line. NC-MiC6 is indicated by arrow and human chromosome 10 by arrowhead. **(a)** FISH using **(i)** neocentromeric probe B153g5 and **(ii)** p-arm probe B326h7 (green), showing the transfer of MC-MiC6 into 293trex. **(b)** Split images of (a) showing DAPI staining.

Figure 7 is a photomicrographic representation showing FISH analysis NC-MiC6 in HCT116pgrxr **(a)** and 293trex **(b)**. NC-MiC6 is indicated by arrow and chromosome 10 by arrowhead. **(i)** FISH using B513g5 NC probe (green) and α-satellite DNA pTRA7 (red). **(ii, iii)** split images for pTRA7 and DAPI, respectively, showing absence of centromere-specific α-satellite DNA on NC-MiC6.

Figure 8 is a photomicrographic representation showing FISH analysis of NC-MiCs 53g, 8a and 20f. NC-MiCs are indicated by arrow. **(a)** Combined FISH images using B153g5 NC cosmid probe (green) and mouse major satellite DNA probe (red), showing absence of major satellite on NC-MiC53g **(i)**, NC-MiC8a **(ii)** and NC-MiC20f **(iii)**. **(b)** Split images of (a) showing DAPI staining.

Figure 9 is a photomicrographic representation showing FISH analysis of NC-MiCs 53g, 8a and 20f. NC-MiCs are indicated by arrow. **(a)** Combined FISH images using B153g5 NC cosmid probe (green) and mouse major satellite DNA probe (red), showing absence of minor satellite on NC-MiC53g **(i)**, NC-MiC8a **(ii)** and NC-MiC20f **(iii)**. **(b)** Split images of (a) showing DAPI staining.

Figure 10 is a photomicrographic representation showing FISH analysis of NC-MiCs 53g, 8a and 20f. NC-MiCs are indicated by arrow. **(a)** Combined FISH images using B153g5 NC cosmid probe (green) and mouse cot DNA probe (red), showing absence of mouse DNA on NC-MiC53g **(i)**, NC-MiC8a **(ii)** and NC-MiC20f **(iii)**. **(b)** Split images of (a) showing DAPI staining.

Figure 11 is a photomicrographic representation showing FISH analysis of NC-MiCs 53g, 8a and 20f. NC-MiCs are indicated by arrow. **(a)** FISH using zeocin resistance gene (green) showing presence of zeocin resistance gene on NC-MiC53g **(i)**, NC-MiC8a **(ii)** and NC-MiC20f **(iii)**. **(b)** Split images of (a) showing DAPI staining.

Figure 12 is a photomicrographic representation showing FISH analysis of tissues derived from chimeric mice PL, CH and KM. Mardel(10) is indicated by arrow. **(a)** Combined

image of FISH using a 10q25 neocentromere-specific E8 probe (green) and human cot DNA (red), showing presence of mardel(10) in cells cultured from PL's lung **(i)**, PL's spleen **(ii)**, CH's tail **(iii)** and KM's tail **(iv)**. **(b)** FISH using E8 NC probe (green) and mouse cot DNA (red), showing absence of mouse DNA on mardel(10) in cells established from PL's lung **(i)**, PL's spleen **(ii)**, CH's skin **(iii)** and KM's tail **(iv)**.

Figure 13 is a photomicrographic representation showing FISH analysis of tissues derived from chimeric mice PL, CH and KM. Mardel(10) is indicated by arrow. **(i-iii)** Combined image of FISH using E8 probe (red) and human chromosome 10 paint (green), showing positive painting only on mardel(10) in cell cultures established from PL's skin **(i)**, CH's tail **(ii)** and KM's tail **(iii)**.

Figure 14 is a photographic representation showing expression of gene hCG40944 in mouse tissue. Quantitative RT-PCR using SYBR green was carried out using primer pair F1/R1 on RNA isolated from chimeric mouse DT, and wild type (WT), tissue. 15 µl of the final reaction mix was run on a 1% w/v agarose gel. Lane M contains DNA size markers. Note gene expression in the brain (but not liver) of mardel(10)-containing chimeric mouse DT but not in the brain of wild type mouse.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is predicated in part on the determination that a region of DNA corresponding to centromeric or neocentromeric DNA is transcriptionally active. This permits the development of artificial or engineered chromosomes and site directed insertion of genetic material to centromeric or neocentromeric regions or immediately adjoining regions or proximal regions of eukaryotic cells such as of mammalian, avian, plant, or other higher eukaryotic origin. Such genetic material surprisingly remains transcriptionally active.

Accordingly, one aspect of the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence corresponding to a centromeric or neocentromeric region of mammalian, avian, plant, or other higher eukaryote DNA, said nucleic acid molecule comprising a heterologous nucleic acid molecule inserted within said centromeric or neocentromeric region or immediately adjoining or proximal region and which heterologous nucleic acid molecule is expressible or otherwise imparts a phenotypically observable effect on a cell carrying the heterologous nucleic acid molecule or on an organism or plant comprising said cell.

Generally, the subject nucleic acid molecule is a DNA molecule. In one form, the DNA molecule is in isolated form. In another form, the DNA is resident within the cell of the mammalian, avian species or plant or any other higher eukaryote. The term "resident" includes the DNA existing as a self-replicating unit relative to the cell's chromosome as well as being integrated into the cell's chromosome.

The term "mammal" includes a human or other primate, a livestock animal (e.g. sheep, cow, pig, horse, donkey, goat), a laboratory test animal (e.g. mouse, rat, rabbit, guinea pig, hamster), a companion animal (e.g. dog, cat) or captive wild animal. An avian species includes a poultry bird (e.g. chicken, duck, turkey, goose), game bird (e.g. wild duck, pheasant, peacock, emu, ostrich) or caged or aviary bird (e.g. parrots, pidgeons, friches). A plant may be a monocotyledonous or dicotyledonous plant, wooded or non-wooded plant,

crop or ornamental plant.

Preferably, however, the DNA is present in a mammalian cell and even more preferably, a human cell.

Accordingly, another aspect of the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence corresponding to a centromeric or neocentromeric region of mammalian DNA, said nucleic acid molecule comprising a heterologous nucleic acid molecule inserted within said centromeric or neocentromeric region or immediately adjoining or proximal region and which heterologous nucleic acid molecule is expressible or otherwise imparts a phenotypically observable effect on a cell carrying the heterologous nucleic acid molecule or on an organism or plant comprising said cell.

Reference herein to a "heterologous gene" means a gene not generally resident within the centromeric or neocentromeric DNA or immediately adjoining or proximal DNA. The term "gene" is used in its broadest sense to include a genomic gene (including exon or intron DNA) as well as cDNA (generally only exon DNA). However, the present invention extends to the incorporation of intronic DNA which, upon transcription and optional splicing, is involved in genetic networking.

The present invention further extends to the use of genetic material to inactivate or activate a gene or genetic locus within a centromeric or neocentromeric region or immediately adjoining or proximal region. In one embodiment, the genetic material induces RNAi or antisense RNA.

The term "gene" should not be construed as limiting the inserted nucleotide sequence to encoding a proteinaceous product as the nucleotide sequence may encode an RNA molecule or a sense molecule or may induce RNAi which is involved in co-suppression or post-transcriptional or translational gene silencing or an intron involved in genetic networking.

By "genetic networking" is meant the modulation of expression of genes, promoters, regulatory regions and peptides, polypeptides or proteins within the genome or proteome of a cell.

Reference to a "centromere" or "neocentromere" includes reference to a functional neocentromere or a functional derivative thereof or a latent, synthetic or hybrid form thereof and is capable of facilitating sister chromatid cohesion and chromosomal segregation during mitotic cell divisions and/or is capable of associating with CENP-A and/or CENP-C and/or other functionally important centromere proteins and/or is capable of interacting with anti-CENP-A antibodies or anti-CENP-C antibodies or antibodies to other functionally important centromere proteins. Generally, and preferably, the neocentromere is incapable of interacting with CENP-B or anti-CENP-B antibodies. Alternatively, the neocentromere may be a latent centromere capable of activation by epigenetic mechanisms or other relevant mechanisms including chromatin reorganization. The neocentromere may also be a hybrid or other human, mammalian, plant, yeast or eukaryote neocentromeres. Synthetic or artificial or engineered neocentromeres provided by, for example, polymeric techniques to arrive at the correct conformation are also contemplated by the present invention. All such forms and definitions of neocentromeres are encompassed by use of this term.

In particular, the centromeric/neocentromeric region is defined at least in humans as within a 4-Mb genetic region, but not limited to this size, encompassed by S/MAR and comprising S/MAR, CENP-H, CENP-A, HP1α, and other essential centromere proteins. A summary of genes expressed in this region is provided in Table 1.

Accordingly, in a preferred embodiment, the present invention is directed to an isolated nucleic acid molecule comprising a nucleotide sequence corresponding to a centromeric or neocentromeric region of human DNA, said nucleic acid molecule comprising a heterologous nucleic acid molecule inserted within said centromeric or neocentromeric region or immediately adjoining or proximal region and which heterologous nucleic acid

molecule is expressible or otherwise imparts a phenotypically observable effect on a cell carrying the heterologous nucleic acid molecule or on an organism or plant comprising said cell, wherein the centromeric or neocentromeric region comprises a q and p arm domain, CENP-H, HP1 domain and a scaffold domain and comprises a gene selected from but not limited to hCG41809, hCG40976, hCG1781464, hCG39839, hCG1781461, hCG40945 and hCG1818126 (see Table 1).

An equivalent region in other mammalian, avian species, plants and higher eukaryotic organisms is also contemplated by the present invention.

Furthermore, the present invention enables the generation of artificial or engineered chromosomes carrying heterologous genes or other genetic material for use in modifying the genotype or phenotype of a cell or higher organism or plant carrying such a cell and/or for use in genetic therapy, genetic improvement, or recombinant protein production.

Accordingly, another aspect of the present invention provides an artificial or engineered chromosome comprising an isolated nucleic acid molecule comprising a nucleotide sequence corresponding to a centromeric or neocentromeric region of mammalian, avian or plant or higher eukaryote DNA, said nucleic acid molecule comprising a heterologous nucleic acid molecule inserted within said centromeric or neocentromeric region or immediately adjoining or proximal region and which heterologous nucleic acid molecule is expressible or otherwise imparts a phenotypically observable effect on a cell carrying the heterologous nucleic acid molecule or on an organism or plant comprising said cell.

More particularly, the present invention contemplates a mammalian artificial or engineered chromosome comprising an isolated nucleic acid molecule comprising a nucleotide sequence corresponding to a centromeric or neocentromeric region or its immediately adjoining or proximal region of mammalian DNA, said nucleic acid molecule comprising a heterologous nucleic acid molecule inserted within said centromeric or neocentromeric region or its immediately adjoining or proximal region and which heterologous nucleic acid molecule is expressible or otherwise imparts a phenotypically observable effect on a

cell comprising the heterologous nucleic acid molecule or on an organism or plant carrying said cell.

Even more particularly, the present invention is directed to a human artificial or engineered chromosome comprising an isolated nucleic acid molecule comprising a nucleotide sequence corresponding to a centromeric or neocentromeric region or its immediately adjoining or proximal region of human DNA, said nucleic acid molecule comprising a heterologous nucleic acid molecule inserted within said centromeric or neocentromeric region or its immediately adjoining or proximal region and which heterologous nucleic acid molecule is expressible or otherwise imparts a phenotypically observable effect on a cell carrying the heterologous nucleic acid molecule or on an organism or plant comprising said cell.

Again, the artificial or engineered chromosomes may be in isolated form or within a cell.

The present invention contemplates an isolated cell or a cell *in situ* comprising an artificial or engineered chromosome or nucleic acid.

Furthermore, the present invention contemplates a method for modifying a phenotype in a eukaryotic cell, said method comprising inserting a genetic sequence, capable of modifying the genome or proteome of the cell, when expressed in said cell into a centromeric or neocentromeric region or its immediately adjoining or proximal region of a chromosome or artificial or engineered chromosome and, in the case of an artificial or engineered chromosome, introducing the artificial or engineered chromosome into a cell.

The present invention provides, therefore, a construct for use in gene therapy, genetic improvement, or recombinant protein production. The construct generally comprises a centromeric or neocentromeric region having a genetic sequence inserted therein, generally operably linked to a promoter and optionally a terminator and/or other regulatory sequences.

Reference herein to a "promoter" is to be taken in its broadest context and includes the transcriptional regulatory sequences of a classical genomic gene, optionally including upstream activating sequences, enhancers and silencers. A promoter is usually, but not necessarily positioned upstream or 5', of a structural gene region, the expression of which it regulates. Furthermore, the regulatory elements comprising a promoter are usually positioned within 2kb of the start site of transcription of the gene.

In the present context, the term "promoter" is also used to describe a synthetic or fusion molecule, or derivative which confers, activates or enhances expression of a nucleic acid molecule in a cell.

Preferred promoters may contain additional copies of one or more specific regulatory elements, to further enhance expression of the sense molecule and/or to alter the spatial expression and/or temporal expression of the sense molecule.

Placing a nucleic acid molecule under the regulatory control of a promoter sequence means positioning the molecule such that expression is controlled by the promoter sequence. As stated above, promoters are generally positioned 5' (upstream) to the genes that they control. In the construction of heterologous promoter/structural gene combinations, it is generally preferred to position the promoter at a distance from the gene transcription start site that is approximately the same as the distance between that promoter and the gene it controls in its natural setting, i.e. the gene from which the promoter is derived. As is known in the art, some variation in this distance can be accommodated without loss of promoter function. Similarly, the preferred positioning of a regulatory sequence element with respect to a heterologous gene to be placed under its control is defined by the positioning of the element in its natural setting, i.e. the genes from which it is derived. Again, as is known in the art, some variations in this distance can also occur.

Examples of promoters suitable for use in the constructs of the present invention include mammalian (e.g. human) viral, fungal, animal and plant derived promoters capable of functioning in plant, animal, insect, fungal or yeast cells. The promoter may regulate the

expression of the structural gene component constitutively, or differentially with respect to cell, the tissue or organ in which expression occurs.

In the present context, the term "operably linked" or similar shall be taken to indicate that expression of the structural gene region or multiple structural gene region is under the control of the promoter sequence with which it is spatially connected in a cell.

In some or many situations, a nucleic acid molecule is under the control of its endogenous promoter where the two molecules are operably linked in their naturally occurring configuration.

Means for introducing (i.e. transfecting or transforming) cells with the constructs are well-known to those skilled in the art.

The constructs described herein are capable of being modified further, for example, by the inclusion of marker nucleotide sequences encoding a detectable marker enzyme or a functional analogue or derivative thereof, to facilitate detection of the synthetic gene in a cell, tissue or organ in which it is expressed. According to this embodiment, the marker nucleotide sequences will be present in a translatable format and expressed, for example, as a fusion polypeptide with the translation product(s) of any one or more of the structural genes or alternatively as a non-fusion polypeptide. The term "structural gene" includes a gene which encodes RNA (e.g. mRNA) or an intronic or exonic RNA.

Genetic constructs are particularly suitable for the transformation of a eukaryotic cell to introduce novel genetic traits thereto or to repair defective genes (i.e. gene therapy). Such additional novel traits may be introduced in a separate genetic construct or, alternatively, on the same genetic construct which comprises the synthetic genes described herein.

The present invention is further described by the following non-limiting Examples.

EXAMPLE 1

Materials and methods

Cell culture

Chinese Hamster Ovary (CHO) cell lines and derivative somatic cell hybrids were cultured as follows. Mouse F9 teratocarcinoma cells and derivatives were cultured in Dulbeccos Modified Eagles Medium (Trace Biosciences) supplemented with 10% v/v FCS, penicillin, streptomycin. Growth medium for hybrid lines containing the zeocin-tagged mardel(10) chromosome (Saffery *et al., Proc. Natl. Acad. Sci. USA 98(10):* 5705-5710, 2001) was supplemented with 200 μg/ml zeocin (Invitrogen). All cells were maintained at sub confluency and were split 1:4 at 24 hr prior to RNA isolation to ensure logarithmic growth at harvest.

Production of somatic cell hybrid lines

Somatic cell hybrid lines 5f and 1f containing mardel(10) or an unrelated normal chromosome 10, respectively, were previously described (du Sart *et al., Nat. Genet. 16(2):* 144-153, 1997). Four new monochromosomal somatic cell hybrid lines were generated using microcell-mediated chromosome transfer (MMCT) procedures (Saffery *et al.* [2001] *supra*). Two of these (CHOM10 and CHON10, respectively) were CHO-based and contained mardel(10) or the progenitor normal chromosome 10 derived from the mardel(10) patient's father as previously described (Barry *et al., Genome Res. 10(6):* 832-838, 2000). These cell lines were produced using selection for the Glutamate Oxaloacetic Transaminase gene as described before (du Sart *et al.* [1997] *supra*). The remaining two hybrid lines (F9M10 and F9N10) were mouse F9 cell-based and contained mardel(10) or normal chromosome 10, respectively. These were produced using the zeocin-tagged mardel(10) under zeocin selection.

RNA extraction and cDNA synthesis

RNA was isolated from cultured cells using Trizol reagent (Life Technologies, Bethesda, NY) or the Qiagen RNeasy midi kit, according to the manufacturer's instructions. RNA levels were quantified by spectrophotometry and integrity of RNA was assayed by non-denaturing gel electrophoresis. Two micrograms of total RNA was used in the production of cDNA using the ABI Reverse transcriptase kit with random hexamer priming according to the manufacturers instructions. One twentieth of this reaction was used in each quantitative RT-PCR reaction.

Primer design and quantitative RT-PCR

Primers for PCR amplification were designed using Primer express software (ABI). Where possible several primer pairs were designed for each gene. To avoid the amplification of contaminating genomic DNA or total RNA, all primer pairs were designed so that at least one of the pair spanned a genomic exon/intron boundary. Each primer was checked for uniqueness in the human genome prior to synthesis. Initial validation experiments were performed for each primer pair to ensure that no amplification was detected from human genomic DNA or total RNA prior to reverse transcription. Quantitative RT-PCR was carried out using SYBR green technology with the Applied Biosystems SYBR green master mix, and reactions were performed on an ABI 7700 Sequence Detection System. Delta CT analysis was used to calculate the relative amount of expression of individual genes in relation to an 18S control amplicon (Ambion Inc.). Further validation experiments were carried out to ensure that the efficiency of amplification of test primer pairs was comparable to that of the 18S control. This involved serial dilution of template (two-fold dilutions to 1/256) followed by PCR amplification with test primers and 18S control. Delta CT for each dilution was then calculated and if efficiencies of amplifications were comparable this value did not change significantly with each dilution. Several primer pairs failed one or more of the validation experiments and were not included in the final analysis. For TAQman-based quantitative RT-PCR, *Assay on demand* pre-optimized

primer and probe mix were employed with TAQman master mix and TAQman 18S control reagents (Applied Biosystems).

Scaffold isolation and array (SIA) analysis

Isolation of cell nuclei:

2×10^8 cells were pelleted at 500 g and washed in PBS for 5 min. Cells were resuspended and washed three times for 5 min at 500 g in isolation buffer containing 3.75 mM Tris-HCl, 0.05 mM Spermine, 0.125 mM Spermidine, 1% v/v thiodyglycol, 20 mM KCl, 0.1 mM PMSF, 0.5 mM EDTA/KOH and 10 KIU/ml Aprotinin [pH 7.4]. Washed cells were resuspended in 12 ml of ice-cold isolation buffer containing 0.1% w/v digitonin and 100 KIU/ml Aprotinin, and broken up in a Dounce type tissue homogeniser with 12 strokes of a B (loose) pestle. Nuclei were collected by three washes in isolation buffer containing 0.1% w/v digitonin and 10 KIU/ml Aprotinin at 900g, 10 min at 4°C. The washed pellet was resuspended in 5 ml isolation buffer containing 0.1% w/v digitonin, 100 KIU/ml Aprotinin and without EDTA/KOH. Nuclei were then filtered through a 40-micron filter to remove nuclei clumps.

Low-salt (LIS) scaffold extraction:

1×10^6 nuclei in 100 ul of isolation buffer with 0.1% w/v digitonin, 100 KIU/ml Aprotinin and without EDTA/KOH were stabilized at 37° C for 20 min. The nuclei were then diluted with 1 ml of LIS buffer consisting of 5 mM Hepes/NaOH, 0.25 mM Spermidine, 2 mM EDTA/KOH, 2 mM KCL and 50 mM 3,5-diiodasalicylic acid, lithium salt (SERVA), and left to extract for 10 min at 4°C. The extracted nuclei were centrifuged at 2,400 g for 20 min at 4°C, followed by washing the pellet four times with 8 ml of digestion buffer consisting 20 mM Tris-HCl, 0.05 mM Spermine, 0.125 mM Spermidine, 20 mM KCl, 0.1 mM PMSF, 0.1% w/v digitonin, 50 mM NaCl, 5 mM $MgCl_2$ and 100 KIU/ml Aprotinin. Restriction enzymes *Eco*RI, *Eco*RV and *Bam*HI were then added at 1000 U/ml and incubated at 37°C for 5 hr. The nuclear scaffold attached DNA was pelleted from the digested loop DNA by centrifugation at 2,400 g for 10 min at 4°C.

BAC array analysis:

100 ng of BAC DNA was immobilized onto Hybond N+ nylon membranes in a dot blot format (minifold SRC-96, Schleisher and Schueel, Dassel, Germany). Identical membranes were pre-annealed with 5 ug of salmon sperm DNA, and probed with 1 ug of scaffold-attached or loop DNA from 1f and 5f cells, ^{32}P-labeled by random priming and pre-annealed with 5 ug of human Cot-1 DNA. Hybridization and washing were performed at high stringency (0.1 x SSC/0.1% w/v SDS, 65°C). Results were analyzed by a phosphorimager system (Storm 860 Gel and Blot Imaging System, Molecular Dynamics) using Image QuaNT version 4.2 software (Molecular Dynamics). The signals obtained using the scaffold-attached DNA probe were compared to those on a duplicate blot hybridised with the loop DNA probe. The percentages of scaffold/matrix attachment for individual BAC spots were calculated by dividing the scaffold/matrix-attached signal by the sum of the scaffold/matrix-attached and loop DNA signal. The mean values from 10 independent experiments and standard deviations were plotted graphically using the midpoint for each BAC on the contig map. Statistical significance was determined using a two-tailed heteroscedastic Student's *t*-test.

Scaffold-FISH

Actively growing cells were harvested by mitotic shake off, washed in phosphate-buffered saline (PBS), and resuspended at 2 x 10^6 cells/ml in 0.0075M KCl for 10 min at 37°C. Cells were then washed in ice-cold PA buffer (15 mM Tris-HCl, 0.2 mM Spermine, 0.5 mM Spermidine, 0.5 mM EGTA, 2 mM EDTA, 80 mM KCl, 20 mM NaCl, 0.1 mM $CuSO_4$ [pH7.2]) at 8 x 10^6 cells/ml before being resuspended at 1 x 10^7 cells/ml in cold PA buffer containing 1 mg/ml digitonin. Nuclei were spun out at 200 g for 10 mins at 4°C, and supernatant containing isolated metaphase chromosomes collected. Chromosomes were spread onto slides and allowed to dry for 16 hr. Slides were then gently lowered horizontally into CIB solution (10 mM Tris, 10 mM EDTA, 0.1% Nonidet P-40, 0.1 mM CuSO4, 20 ug/ml PMSF [pH 8.0]) for 5 mins and then extracted in CIB containing 0.5 M NaCl for 5 mins. FISH was carried out using standard conditions.

Chromatin immunoprecipitation and array (CIA) analysis

CIA analysis for CENP-H and HP1 was carried out essentially as described for CENP-A (Lo *et al., EMBO J. 20(8):* 2087-2096, 2001A; Lo *et al., Genome Res. 11(3):* 448-457, 2001B) using an affinity-purified rat anti-CENP-H antibody (Sugata *et al.,* [2000] *supra*), anti-HP1α (Le Douarin *et al., EMBO J. 15(23):* 6701-6715, 1996), and anti-HP1β (Serotec Ltd, Oxford, UK). Micrococcal nuclease digestion of chromatin was carried out using 4 units per mg of chromatin-associated DNA for 5-6 min to obtain polynucleosomes.

EXAMPLE 2

Genes in neocentromeric region

Human neocentromere activation generally occurs in euchromatic regions of the genome containing characterized or predicted genes (Choo[2001] *supra*; Amor and Choo, *Am. J. Hum. Genet. 71(4):* 695-714, 2002). Available human genome sequence databases have been used to identify putative genes in the vicinity of a 10q25 neocentromere on the mardel(10) marker chromosome (du Sart *et al.* [1997] *supra*). A direct comparison of gene predictions in the public database [Ensembl, UCSC genome browser] to those in the Celera database revealed several differences in both gene number and order. Independent mapping experiments supported the Celera gene order at 10q25. Celera annotations were used as the basis for gene identification. Figure 1a shows the arrangement of predicted genes with respect to the previously mapped CENP-A-associated region on mardel(10) (Lo *et al.,* [2001A] *supra*). In Figure 1a, the BAC array spanning a total of 8 Mb showed positions of clones (horizontal bars) used in CIA and SIA analyses. Positions and orientations of genes located at 10q25 used in the expression study are shown by arrows or arrowheads. The location of the CENP-A-associated domain is indicated by purple shading (Lo *et al.* [2001A] *supra*). A total of 51 genes within an 8-Mb region were examined, including a single putative gene (Celera gene ID: hCG39837) that spans the CENP-A-associated domain.

Three sets of somatic cell hybrids containing either mardel(10) or a normal chromosome 10 in comparable genetic backgrounds were used in quantitative real-time PCR using SYBR green for gene expression study both before and after centromere activation (Table 1). Of 51 genes examined, 14 showed expression in one or more somatic cell hybrid lines tested (Figure 2; Table 1; green arrows in Figure 1a). Of these, 8 were examined further using TAQman quantitative RT-PCR, with essentially similar results. Surprisingly, no difference in expression level was detected between corresponding hybrids containing the normal chromosome 10 or mardel(10), indicating that the process of centromere activation that resulted in major remodelling of underlying chromatin (see below) had no measurable effect on the expression of these genes. Genes as close as 200 kb (hCG40949) to the core CENP-A domain were expressed at unaltered levels. Expression of hCG39837, the only gene spanning the CENP-A domain, could not be detected but given that this gene is expressed predominantly in brain (Su *et al., Proc. Natl. Acad. Sci. USA 99*: 4465-4470, 2002), this was not surprising. Therefore, any effect on gene expression following CENP-A deposition into nucleosomes could not be directly measured. Chromosome 10 genes used in this expression analysis are summarized in Table 2.

EXAMPLE 3

Detection of scaffold-attached chromatin domain at 10q25 neocentromere

The absence of any functional domain definition has previously not allowed gene positions to be directly related to centromere activity. The available sequence of the 10q25 neocentromere now provides a unique opportunity to define the relative positions of centromeric chromatin domains and directly assay the effects these domains have on underlying gene expression.

The pattern of chromosomal scaffold/matrix attachment at 10q25 both pre- and post-neocentromere formation was investigated. The chromosomal scaffold is the insoluble chromatin that remains following removal of core histones (Paulson and Laemmli, *Cell 12:* 8178-828, 1977). It interacts directly with DNA through specific S/MAR (or scaffold/matrix attachment region) sequences, and contains proteins such as

Topoisomerase IIα, CENP-C, and cohesin subunits, that have been shown to be essential for centromere function (Pinsky *et al., Dev. Cell. 3(1):* 4-6, 2002, Kalitsis, [1998] *supra*, Saitoh *et al., Bioassays 17(9):* 2919-2926, 1995). A novel scaffold isolation and array (SIA) analysis procedure was developed involving the differential salt extraction of scaffold-attached and non-scaffold-attached (or loop) DNA for use as hybridisation probes on a previously described 10q25 BAC-array (see Example 1). Using this approach, an approximately 3.5-Mb region at the 10q25 neocentromere was identified showing a significantly increased level of chromosomal scaffold/matrix attachment compared to the corresponding region of the normal chromosome 10 (Figure 1b). In Figure 1b, scaffold/matrix attachment along the 10q25 BAC array was determined by SIA analysis on chromatin prepared from 5f [mardel(10) chromosome] and 1f (normal chromosome 10) hybrid cell lines. Data-points, represented on the x-axis by the midpoints of the positions of the BACs relative to the start of the contig map, were expressed as the means and standard deviation of the means from 10 independent experiments and were calculated and are shown on the y-axis as the percentage difference between the scaffold attached/unattached signal ratio of 5f and 1f. Significance of the data-points was determined using a student's *t*-test and is indicated by an asterisk ($p<0.01$). The S/MAR-enriched domain is indicated by blue shading. A summary of scaffold-FISH data is shown at the top of the graph, denoting scaffold-attached (+) or non-scaffold-attached (–) BAC DNA, and showing close concordance with the SIA results. Corresponding BACs on the graph used in FISH analysis are shown as open circles.

This region contains at least 30 putative genes, eight of which were identified as being expressed in one or more of the somatic cell hybrid lines (Example 2). For independent verification of the SIA results, direct visualization of scaffold attachment across the region of interest was performed by FISH analysis of metaphase chromosomes (Bickmore and Oghene, *Cell 84(1):* 95-104, 1996) using as probes BAC clones from the 10q25 array. The results, shown in Figure 3 for several BAC clones and summarized in Figure 1b (+/- symbols), were in tight concordance with those obtained using SIA analysis. FISH on salt-extracted, histone-depleted chromosomes was carried out essentially as previously described (Bickmore and Oghene [1996] *supra*). In Figure 3, panels A and D show FISH

using BAC clones BA313D6 and BA427L15, which mapped outside the S/MAR-enriched domain identified by SIA analysis which produced dispersed signals (open arrows) on both the normal chromosome 10 (top panel) and mardel(10) chromosomes (bottom panel), indicating predominantly non-scaffold attachment of the probed regions. Panels B and C show FISH using BAC clones E8 and BA153G5 mapping within the S/MAR domain which produced dispersed signals (open arrow) on chromosome 10 (top panel) but tightly packed signal on the maardel chromosome (closed arrow; bottom panel), indicating predominantly scaffold attachment of the probed regions on mardel(10). This increase in S/MAR over a substantial region may explain the tighter compaction of chromatin that gives rise to the mardel(10) primary constriction. As shown in Figure 1b, the previously identified CENP-A domain is located centrally within the 3.5 Mb domain of enhanced chromosomal S/MAR-modified chromatin. These results clearly demonstrate the existence of a substantive scaffold-attached chromatin domain at the 10q25 neocentromere, and demonstrate that such a domain and its constituent proteins have no measurable effect on underlying gene expression. The Accession number of BACs used in this study are shown in Table 3.

EXAMPLE 4

Detection of gene expression in CENP-H domain

In view of both the paucity of genes and lack of detectable expression within the CENP-A-binding domain, the effects of the binding of a second essential constitutive centromere protein CENP-H (Sugata *et al.* [1999] *supra*; Sugata, *et al.* [2000] *supra*) have on gene expression was examined. A polyclonal rat anti-human CENP-H antibody was used in a chromatin immunoprecipitation and array (CIA) mapping procedure previously used to define the CENP-A-binding domain (Lo *et al.* [2001A] *supra*). A 900-kb domain of CENP-H association was identified at a site that overlapped minimally with the distal q-arm edge of the scaffold-enriched region (Figures 1b and 1c). These figures show the distribution of CENP-H antigen along the 10q25 BAC contig (x-axis) as determined by CIA analysis. The y-axis shows the fold difference between the normalised bound/input ratio of mardel(10)- and normal chromosome 10-containing cell lines. Each data-point is

the mean of four independent CIA experiments. Significance of the data-points was determined using a student's t-test and is indicated by an asterisk ($p<0.01$). The position of the CENP-H-associated region is indicated by green shading. This CENP-H-binding domain is approximately 1 Mb away from, and showed no overlap with, the CENP-A-associated domain (Figure 1C). The non-overlapping nature of CENP-H and CENP-A domains is of note in light of evidence showing that CENP-C is targeted to CENP-A-containing chromatin, and that both CENP-A and CENP-H are required for CENP-C localization (Howman *et al., Proc. Natl. Acad. Sci. USA 97(3):* 1148-1153, 2000; Van Hooser *et al., J. Cell. Sci. 114(19):* 3529-3542, 2001; Fukagawa *et al., EMBO J. 20(16)*: 144-153, 2001; Ando *et al., Mol. Cell. Biol. 22(7):* 2229-2241, 2002). This suggests the adoption of complex higher-order interactions between these protein domains (Figure 1e). Importantly, 6 genes were identified within this region that were expressed at comparable levels both pre- and post-neocentromere formation, suggesting that the constitutive, cell cycle independent association of an essential centromere protein such as CENP-H is not inhibitory to underlying gene expression.

EXAMPLE 5

Relationship between heterochromatin protein and centromere function

The specific role of heterochromatin and its associated proteins at the centromeric/pericentromeric regions remains unclear. Numerous studies have indicated a role in gene silencing although several genes have been described that escape this repression presumably through some insulating activities protecting the genes from heterochromatin protein encroachment (Schulze *et al., Mol. Gen. Genet. 264(6):* 782-789, 2001). Other studies have also suggested that heterochromatin may play a role in sister-chromatid cohesion (Vagnarelli *et al., Chromsoma 110(6):* 393-401, 2001; Bernard *et al., Science 294(5551):* 2539-2542, 2001; Bernard and Allshire, *Trends Cell Biol. 12(9):* 419, 2002). However, the existence of an abundance of repetitive DNA at and around the centromeres used in these studies has hindered the functional dissection of any direct role heterochromatin may have in kinetochore activity. The binding of heterochromatin protein HP1 at neocentromeres in the absence of centromeric/pericentromeric repetitive DNA

(Saffery *et al., Hum. Mol. Genet. 9(2):* 175-185, 2000) strongly suggests a direct role of heterochromatin in mammalian centromere function.

In order to investigate the relationship between heterochromatinisation and transcriptional activity at the neocentromere, a polyclonal anti-HP1α antibody (Le Douarin *et al.* [1996] *supra*) was used in CIA experiments to determine the extent and position of HP1 association. A domain of approximately 100 kb was identified which is significantly enriched for HP1 (Figure 1d). Figure 1d shows the distribution of HP1α antigen along the 10q25 BAC contig (x-axis) as determined by CIA analysis. The y-axis shows the fold difference between the normalised bound/input ratio of mardel(10)- and normal chromosome 10-containing cell lines. Each data-point is the mean of four independent CIA experiments. Significance of the data-points was determined using a student's *t*-test and is indicated by an asterisk ($p<0.01$). The position of the HP1α-associated region is indicated by orange shading. A similar result was obtained using a second HP1 antibody to a β isoform. The HP1-binding domain maps approximately 800 kb from the core CENP-A region on the p-arm side of mardel(10). Between the HP1 and CENP-A domains is a single expressed gene (GFRA1; Table 1) that is not inhibited following neocentromere activation. The 100-kb HP1 domain itself encompasses a single gene (PNLIP) that was not expressed in any of the cell lines tested. This is not surprising given that this gene shows a pancreas-specific expression profile (Su *et al.* [2002] *supra*). Therefore, as with the CENP-A domain, any direct effect on gene expression following HP1 binding could not be directly determined and will await future analysis in another more appropriate cell type.

EXAMPLE 6

Human cell models for NC-MiC analysis and gene expression

Truncation of mardel(10) in human fibrosarcoma HT1080 cell line was performed *via* transfection of constructs containing a targeting DNA, human telomere sequence, and a hygromycin-resistance selection marker (Saffery *et al.* [2001] *supra*). Recent approaches have involved the use of a similar construct with a neomycin resistance gene (neo^R) with flanking loxP sites, located between the p-arm-targeting DNA and human telomere

sequence. This construct was transfected into a HT1080 cell line carrying a 16 Mb NC-MiC2 (Saffery *et al.* [2001] *supra*). Screening of 15,000 colonies had produced several second-generation NC-MiCs containing further truncation of the arm of NC-MiC2, and incorporating the loxP sequence (Wong *et al., Gene Ther 9:* 724-726, 2002). One of these, a linear 1.2 Mb NC-MiC6, resulted from a targeted truncation event as evidenced by pulsed field gel electrophoresis, PCR, and FISH analysis (Figure 4b) (Wong *et al.* [2002] *supra*). NC-MiC6 shows full mitotic stability and centromere protein-binding properties, thereby demonstrating normal neocentromere function. Importantly, it contains loxP sites that can be used for later insertion of genes *via* CRE-mediated recombination.

NC-MiC6 was successfully transferred into two other human cell lines: human colorectal HCT116pgrxr and human embryonic kidney 293trex cell lines (Figures 5-7). NC-MiC transfer was achieved using MMCT with hygromycin and neomycin selection for clones in the two cell lines respectively. The successful transfer of NC-MiC6 into these cell lines indicated that both the hygromycin and neomycin genes were expressed on NC-MiC6.

The HCT116 cell line used for fusion transfer of NC-MiC6 expresses the insect ecdysone receptor (pgRXR) and carries a zeocin resistant gene. This cell line can be used to express inducible levels of any desired protein. The ecdysone inducible system utilises a dimer of the ecdysone receptor (VgEcR) and the retinoid X receptor (RXR) that binds to a hybrid ecdysone response element (E/GRE) in the presence of ecdysone analog, muristerone. The ecdysone receptor is also modified to contain the VP16 transactivation domain that is derived from *Drosophila*. The addition of muristerone induces the binding of the dimer of RXR and VgEcR to the hybrid Ecdysone response element (E/GRE) which consists of both the natural ecdysone response element and glucocorticoid response element, hence leading to an induction in the expression of the gene of interest.

The 293T cell line expresses Tet repressor (tetR) protein and is resistant to blasdicidin (BsdR). This cell line can also be used to express inducible levels of any protein of interest. In the absence of tetracycline, Tet repressor forms homodimers that bind to Tet operator sequences in the inducible expression vector, repressing transcription of the gene.

Upon addition of tetracycline, tet binds to tetR homodimer, causing the release of the tetR from the operator due to a change in its conformation, thus induction of transcription from the desired gene.

These results illustrate, first, the ease with which NC-MiCs containing a selectable marker can be transferred from one cell line to another using antibiotic resistance genes, second, the degree of mitotic stability of NC-MiC6 in various cell lines, and third, the expression of the selection markers on the NC-MiC. These results combined point to the feasibility of using human cell lines as a model for gene expression from NC-MiCs and, therefore, the future correction of gene defects in human cell model systems.

EXAMPLE 7

NC-MiC production and gene expression in mouse embryonic stem cells

In addition to studies describe above, a mardel(10) chromosome tagged with zeocin-resistance gene was successfully transferred into mouse embryonic stem (ES) cells using MMCT. Truncation constructs as shown in Figure 4a was also used to produce NC-MiCs in these cells. The truncation constructs (Figure 4a) were designed to target specific B43all and B79e16 sites (Figure 4b). Although no targeted events were obtained, four clones of random truncation were generated (Figure 4b), three of which were described further (Figures 8-11). FISH analysis with various probes showed that both truncation and deletion had occurred on the q arms of NC-MiC53g, NC-MiC8a and NC-MiC20f, however, the core CENP-A binding domain (upon which a functional centromere is assembled) remained intact as indicated by the FISH analysis using B153g5 probe. Results of FISH characterization of these minichromosomes (not shown) are summarized in Figure 4b.

Neither major nor minor satellite DNA were found to be present on NC-MiC53g, NC-MiC8a and NC-MiC20f (Figures 8 and 9), showing that these NC-MiCs had not acquired these satellite DNAs following the truncations and that the 10q25 neocentromere was mitotically functional. Moreover, mouse DNA, as shown in FISH analysis using mouse cot DNA (Figure 10), was absent on these NC-MiCs, indicating no detectable integration of

mouse DNA into the NC-MiCs. The positive signals of zeocin resistance gene, previously shown to be present on distal q region (Saffery *et al.* [2001] *supra*), on all three NC-MiCs as shown by FISH (Figure 11) indicated that the zeocin resistance gene has integrated into these NC-MiCs during the truncation process. The presence of this selectable marker will facilitate the future transfer of NC-MiCs into other mammalian cell lines.

An analysis of gene expression from NC-MiCs in ES lines revealed a diverse pattern of expression with some NC-MiC cell lines expressing few of the 14 genes tested (e.g. lines 20fC94, 8aC94, 53g43A expressing hCG40964, hCG1818126; Table 4), while other NC-MiC-containing lines expressing many different genes (e.g. 1.931b expressing 5 different genes; Table 4). The diversity in gene expression mirrors the different human chromosome 10 regions contained in the NC-MiCs within ES cells and supports the use of a mouse ES model for analysis of gene expression from NC-MiCs.

EXAMPLE 8

Establishment of an animal model for gene expression from NCCs and NC-MiCs

The ability of human neocentromeres to function mitotically in whole animal was tested. This was done by studying the behavior of mardel(10), a human neocentromere-containing chromosome (NCC), in mice. Following the transfer of mardel(10)-containing ES cells into mouse blastocysts and reimplantation into foster female mice, a number of high-grade chimeric mice (e.g. PL, CH, KM, DT) were obtained. Analysis of tissue samples from adult animals by PCR using mardel(10)-specific primers demonstrated maintenance of the mardel(10) chromosome in many different mouse tissues including bone marrow, skin, kidney, spleen, caecum, heart, lung, liver, brain and testis, in chimeric mice (Table 5). FISH analysis using mardel(10) probe E8 and human cot DNA confirmed the presence of mardel(10) in several different tissues in three of the mice (PL, CH, KM - examples shown in Figure 12a). Furthermore, the absence of mouse paint (Figure 12b) on mardel(10) and the positive painting of human chromosome 10 paint (Figure 13) in these tissues clearly indicated that there was no detectable integration or rearrangement of DNA between

mardel(10) and the mouse genome. These results illustrate that the mardel(10) neocentromere functions correctly in mice, suggesting that the species barrier will not pose a problem when considering mouse as a model system for studying NC-MiCs in a whole animal model.

In addition, we were able to detect expression of several genes in tissues obtained from several of these chimeric animals (see Table 4). Expression of human genes hCG40949, hCG1811159, hCG1818126, hCG40945, and hCG40995 (Celera ID numbers) was detected in primary tail fibroblasts in two of the animals and expression of hCG40944 (Celera ID number) in RNA isolated from the brain of one of the chimeric animals (DT - see Figure 14) was also demonstrated. This shows the feasibility of a mouse model system for analysing gene expression from neocentromere containing chromosomes (including minichromosomes) in mice and, in addition to the observed mitotic stability of NCC in whole animals, opens the way for using NC-MiCs containing therapeutic genes to correct mouse models of human diseases.

TABLE 1

10q25 gene expression analysis

Gene no. (Fig. 1a)	Celera Gene ID (Gene Name)	Domain	Relative expression levels $2^{-\Delta\Delta CT}$			
			1f (N10) v 5f (M10)	CHON10 v CHOM10	F9N10 v F9M10	Pooled N10 v M10
1	hCG41809	q arm	1.02±0.09	1.14±0.42	1.24±0.38	1.12±0.32 (p=0.98)
2	hCG40976 (hypothetical protein FLJ21952)	q arm	1.33±0.36	1.01±0.22	NE	1.16±0.33 (p=0.95)
3	hCG1811152	q arm	1.09±0.22	0.88±0.27	0.96±0.09	0.99±0.24 (p=0.94)
4.	hCG1781464 (caspase 7 - CASP7)	CENP-H	1.03±0.45	1.4±0.33	1.02±0.58	1.15±0.54 (p=0.87)
5	hCG40995	CENP-H	1.26±0.62	1.39±1.09	1.62±1.53	1.18±0.75 (p=0.84)
6.	hCG39839 (adrenergic β-1 receptor - ADRB1)	CENP-H	1.61±0.93	0.9±0.61	NE	1.19±0.1 (p=0.95)
7.	hCG1781461 (hypothetical protein FLJ10188)	Scaffold CENP-H	0.68±0.24	1.39±0.89	NE	1.05±0.48 (p=0.65)
8.	hCG40945 (tudor domain protein - TRD1)	Scaffold CENP-H	NE	NE	0.83±0.77	0.83±0.77 (p=0.19)
9.	hCG1818126	Scaffold CENP-H	0.70±0.31	1.7±0.88	0.87±0.19	1.04±0.67 (p=0.91)
10.	hCG1811159 (actin binding LIM protein - ABLIM)	Scaffold	1.16±0.84	1.04±0.54	1.64±1.31	1.25±0.92 (p=0.86)
11.	hCG40944	Scaffold	0.71±0.34	1.23±0.74	1.23±0.82	1.01±0.65 (p=0.75)
12.	hCG40949 (tRNA pseudouridine synthase – TRUB1)	Scaffold	1.16±0.65	1.13±0.97	1.9±1.3	1.51±0.98 (p=0.68)
13.	hCG40963 (GDNF family receptor - GFRA1)	Scaffold	1.2±0.61	0.94±0.4	1.24±0.84	1.17±0.89 (p=0.97)
14.	hCG40964	Scaffold	1.03±0.89	1.23±0.83	2.66±2.38	1.45±1.42 (p=0.99)

Gene number refers to Figure 1. Celera Gene ID and corresponding gene name (where applicable) are shown. Domain refers to the location of the gene in relation to domains of chromatin modification identified in the current study. The relative expression levels of genes are shown for somatic

hybrid cell lines in different genetic backgrounds (see Example 1). Relative expression levels were calculated as follows. The C_T value for a particular gene amplification within a particular RNA sample refers to the amplification cycle at which fluorescence exceeds a particular threshold level. This is directly related to the quantity of a particular RNA within the starting sample. CT(gene of interest) - CT(18s rRNA)] are used to calculate ΔΔCT values [ΔΔCT = ΔCT(cell line containing normal chromosome 10) - ΔCT(cell line)]. Relative expression levels ($2^{-\Delta\Delta CT}$) are a direct comparison of expression level from the normal chromosome 10 relative to expression level from the neocentromere region of the mardel(10) chromosome. Statistical significance of any difference in expression level between cell lines was calculated using ΔCT values in a Student's *t*-test. Pairs of hybrids [mardel(10) *versus* normal 10] in identical genetic backgrounds were used for each analysis. Total N10 *versus* M10 refers to a pooled analysis of all ΔCT values for each gene in each of the three hybrid pairs combined. Any statistically significant differences are indicated by asterisks.

TABLE 2

Chromosome 10 genes used in expression analysis

	Celera Gene ID	Gene Name	Chromosome 10 position	Primers used in amplification*
1	hCG41809		112,645,719 112,664,905	F1 5' – GCTGCAGCAAAAAGACCAGAA [SEQ ID NO:1] R1 5' – AGTGGCCCTTGCTTTGGAA [SEQ ID NO:2]
2	hCG40975		113,898,330 113,928,906	
3	hCG1811160		114,053,329 114,110,416	
4	hCG40976	hypothetical protein FLJ21952	114,184,667 114,201,475	F1 5' – TGAGGTGTCCATGACGGAGTCA [SEQ ID NO:3] R1 5' – AAACCCAACCTCAGTGTGGTCC [SEQ ID NO:4]
5	hCG1811152		114,201,402 114,569,789	F1 5' – ACCGAGCAAATTGGTCAGGA [SEQ ID NO:5] R1 5' – TGACCACGATGATCCCTAGGA [SEQ ID NO:6]
6	hCG1776259	hypothetical protein FLJ23556	114,825,469 114,827,846	
7	hCG40993	hyaluronan binding protein 2	115,307,412 115,343,988	
8	hCG1781477		115,344,053 115,346,689	
9	hCG1781473		115,351,473 115,364,947	
10	hCG1811154	nebulin-related anchoring protein, NRAP	115,366,537 115,418,409	
11	hCG1781464	caspase 7 CASP7	115,446,485 115,485,259	F2 5' – CATGCGATCCATCAAGACCA [SEQ ID NO:7] R2 5' – TTTCGAACGCCCATACCTG [SEQ ID NO:8]
12	hCG40992	hypothetical protein FLJ23537	115,526,475 115,537,661	
13	hCG40995		115,588,926 115,608,290	F1 5' – AAATACCTGGAACCGGCTTTAC [SEQ ID NO:9] R1 5' – ATTCAGTGTCCAGTGGCAATG [SEQ ID NO:10]
14	hCG40996		115,609,033	

	Celera Gene ID	Gene Name	Chromosome 10 position	Primers used in amplification*
			115,665,581	
15	hCG1781466	hypothetical protein FLJ20147	115,668,412 115,671,618	
16	hCG39839	adrenergic, beta-1-, receptor ADRB1	115,798,151 115,801,328	F1 5' – AGCTCATCTTTGTGGAGAAGGA [SEQ ID NO:11] R1 5' - CAAGGAACATCAGCAAGCCAC [SEQ ID NO:12]
17	hCG1781461	hypothetical protein FLJ10188	115,880,285 115,887,568	F1 5' – GCAGCTTCAAAGAGGTAAGCA [SEQ ID NO:13] R1 5' – GGATTCAGACTGAAGCTGTGCA [SEQ ID NO:14]
18	hCG1818126		115,891,387 115,913,027	F1 5' – GGATGGAACAGGCCAACAAGA [SEQ ID NO:15] R1 5' – TTCATACAGCTGGTGCAACC [SEQ ID NO:16]
19	hCG1781474		115,918,054 115,929,626	
20	hCG1645882		115,929,623 115,930,005	
21	hCG1781475		115,935,003 115,943,578	
22	hCG40945	tudor domain containing 1 TDRD1	115,958,916 115,983,847	F1 5' – GCAGCTTCAAAGAGGTAAGCA [SEQ ID NO:17] R1 5' - GCACGGTACCACTGATCATCC [SEQ ID NO:18]
23	hCG39838		116,001,411 116,046,696	
24	hCG40946		116,049,975 116,160,243	
25	hCG1811159		116,186,618 116,413,721	F1 5' – AAGGATTTAGCAGCCATTCCG [SEQ ID NO:19] R1 5' – TGGTACCCTTCTGCTGATGGA [SEQ ID NO:20]
26	hCG1640109		116,445,328 116,446,075	
27	hCG40944		116,598,313 116,655,222	F1 5' – GGCTGCAAAGTGCCTTACACA [SEQ ID NO:21] R1 5' - CCAAGCCCCAGTTAATTGCTT [SEQ ID NO:22]
28	hCG40949	tRNA Pseudouridine sythetase TRUB1	116,693,129 116,731,333	F3 5' – AGCCCGAGGAGTTCTGGTTGTT [SEQ ID NO:23] R3 5' – TTTCCCCAGTTCTCCAATGGC [SEQ ID NO:24]
29	hCG39837		116,848,448	

	Celera Gene ID	Gene Name	Chromosome 10 position	Primers used in amplification*
			117,704,615	
30	hCG40963	GDNF family receptor alpha 1, GFRA1	117,819,423 118,028,987	F3 5' – AGATCTCGCCTTGCGGATTT [SEQ ID NO:25] R3 5' – ATGACTGTGCCAATAAGCCCC [SEQ ID NO:26]
31	HcG40968		118,079,885 118,113,360	
32	hCG1790208		118,164,325 118,172,434	
33	hCG1658018		118,182,987 118,232,863	
34	hCG1644138		118,194,235 118,195,098	
35	hCG1640542	pancreatic lipase PNLIP	118,300,940 118,322,397	
36	hCG40961	pancreatic lipase-related protein 1, PNLIPRP1	118,345,556 118,364,209	
37	hCG39836	pancreatic lipase-related protein 2, PNLIPRP2	118,376,340 118,400,160	
38	hCG40969		118,418,814 118,424,987	
39	hCG40967		118,426,195 118,497,629	
40	hCG1781511		118,445,814 118,448,155	
41	hCG1792255		118,504,937 118,505,731	
42	hCG1795809		118,615,272 118,637,149	
43	hCG1640649	KIAA1598 protein KIAA1598	118,639,837 118,661,756	
44	hCG40964		118,666,847 118,760,395	F1 5' – TATTGCTTGCTCCTTCAGACTG [SEQ ID NO:27] R1 5' – CTCCCTCTTTCCCTTTTATTCC [SEQ ID NO:28]

	Celera Gene ID	Gene Name	Chromosome 10 position	Primers used in amplification*
45	hCG40915		118,889,263 118,893,289	
46	hCG39783	solute carrier family 18 SLC18A2	118,997,278 119,036,834	
47	hCG40911		119,038,040 119,131,354	
48	hCG1818116		119,246,337 119,250,541	
49	hCG40913	empty spiracles homolog 2 EMX2	119,298,966 119,305,74	
50	hCG40912	KIAA0941 protein Rab11-FIP2	119,760,468 119,802,156	
51	hCG41145	hypothetical protein FLJ13188	120,066,029 120,099,300	

TABLE 3

Accession number of BACs used in this study

BAC (no.)	BAC NAME	GENBANK ACCESSION NUMBER
1	bA313D6	AL136368
2	bA113F6	AL392167
3	bA324O2	AL391986
4	bA258A12	AL135792
5	bK8502	AL357042
6	bK1031G15	AC006097
7	bA190F19	AL390197
8	bA240G16	AC005886
9	bA86E10	AL355543
10	bA411P18	AL35543
11	bA332D23	AC005383
12	bA108K1	AL354873
13	bA206G17	not available
14	E8	AF222854
15	bA87P3	AC016042
16	bA153G5	AL357059
17	bA5239H22	AL356100
18	bA87E14	AL159173
19	bA48L24	not available
20	bA69K10	not available
21	bA326H7	AC005384
22	bA96n16	AC073588
23	bK1028c12	AC006095
24	bK1137l11	AC005872
25	bA506P9	AC012470
26	bA295o23	AC011328
27	bA33D13	AC011329
28	bA498B4	AC016825
29	bK163G10	AC005660
30	bK bk bK54O2	AC005661
31	bA539i5	AC023283
32	bA501J20	AC022283
33	bK11O6	AC005658
34	bA389e6	AL359836
35	bA328K15	AL139121
36	bK287C20	not available
37	bA319i23	AL365498
38	bA427l15	AL139407
39	bA79a18	not available

TABLE 4

Gene expression from derivatives in mouse

	W9.5	ES20A [mardel (10)]	20fC94	8aC94	53g43a	1.9-31bc43a	DT-Brain	CH-FIB	KM-FIB	CHO-N10
hCG41809							-			+
hCG40976							-			+
hCG1811152							-			+
hCG1781464	-	-?	-	-	-	-	-			+
hCG40995	-	+	-	-	-	+	-	+	+/-	+
hCG39839							-			+
hCG1781461[&]							-	?	?	+
hCG40945	-	+	-	-	-	+	-	+/-	+	-
hCG1818126[*]	-	+	+?	+?	+	+	-	+	+	+
hCG1811159[$]	-	+	-	-	-	+	-	+	+	+
hCG40944	-	+	-	-	-	-	+	?	?	+
hCG40949	-	+	-	-	-	+	-	+	+	+
hCG40963	-	-	-	-	-	+/-	-	?	?	+
hCG40964	-	+	+	+	+	-	-	-	-	+
hCG41421	-	+	-	-	-	-	-	-	-	+

* hCG40950
& hCG40948
$ hCG40947

W9.5 is wild-type embryonic stem (ES) cell line. ES20A is ES cells containing an intact mardel(10). Refer to Figure 4 for explanation of 20fC94, 8aC94, 53g43A, and 1.931b. The brain of chimeric mouse DT and fibroblasts of chimeric mice CH and KM were analyzed. CHO-N10 is a CHO cell line containing a normal human chromosome 10.

TABLE 5

Screening of various tissues for presence of mardel(10) in chimeric mice PL, CH and DT by PCR using specific primers to the mardel(10) marker chromosome

Tissues	PCR results			Tissues	PCR results		
	PL	CH	DT		PL	CH	DT
left lung	+ve	+ve	+ve	adrenal	-ve	+ve	nd
right lung	-ve	-ve	+ve	uterus	-ve	+ve	nd
left kidney	+ve	+ve	nd	ovary	-ve	-ve	nd
right kidney	-ve	+ve	nd	tongue	-ve	-ve	nd
Thymus	-ve	-ve	nd	oesophagus	-ve	-ve	nd
caecum	+ve	+ve	nd	salivery gland	-ve	-ve	nd
small intestine	-ve	-ve	nd	liver	+ve	+ve	-ve
large intestine	-ve	-ve	nd	bone marrow	+ve	+ve	nd
heart	+ve	+ve	nd	spleen	+ve	+ve	nd
brain	+ve	-ve	+ve	Pancreas	-ve	+ve	nd
stomach	-ve	-ve	nd	Tail	-ve	+ve	+ve
pale back skin	+ve	+ve	nd	left leg	-ve	-ve	nd
sternum	+ve	-ve	nd	right leg	-ve	-ve	nd
epididymus	nd	nd	+ve	Testis	nd	nd	+ve

Those skilled in the art will appreciate that the invention described herein is susceptible to variations and modifications other than those specifically described. It is to be understood that the invention includes all such variations and modifications. The invention also includes all of the steps, features, compositions and compounds referred to or indicated in this specification, individually or collectively, and any and all combinations of any two or more of said steps or features.

BIBLIOGRAPHY

Amor and Choo, *Am. J. Hum. Genet. 71(4):* 695-714, 2002.

Ando *et al., Mol. Cell. Biol. 22(7):* 2229-2241, 2002.

Barry *et al., Genome Res. 10(6):* 832-838, 2000.

Bernard and Allshire, *Trends Cell Biol. 12(9):* 419, 2002.

Bernard *et al., Science 294(5551):* 2539-2542, 2001.

Bernat *et al., J. Cell. Biol. 111:* 1519-1533, 1990.

Bickmore and Oghene, *Cell 84(1):* 95-104, 1996.

Brown *et al., Hum. Mol. Genet. 3:* 1227-1237, 1994.

Choo, *Dev. Cell 1(2):* 165-177, 2001.

Clarke and Carbon, *Annu. Rev. Genet. 19:* 29-56, 1995.

du Sart *et al., Nat. Genet. 16(2):* 144-153, 1997.

Earnshaw and Mackay, *FASEB J. 8:* 947-956, 1994.

Earnshaw *et al., Chromosoma 98:* 1-12, 1989.

Farr *et al., EMBO Journal 14:* 5444-5454, 1995.

Fukagawa *et al., EMBO J. 20(16)*: 144-153, 2001.

Grady *et al., Proc. Natl. Acad. Sci. USA 89:* 1695-1699, 1992.

Haaf *et al., Cell 70:* 681-696, 1992.

Howman *et al., Proc. Natl. Acad. Sci. USA 97(3):* 1148-1153, 2000.

Hudson *et al., J. Cell Biol. 141:* 309-319, 1998.

Kalitsis, *Proc. Natl. Acad. Sci. USA 95(3):* 1136-1141, 1998.

Larin *et al., Hum. Mol. Genet. 3:* 689-695, 1994.

Le Douarin *et al., EMBO J. 15(23):* 6701-6715, 1996.

Lo *et al., EMBO J. 20(8):* 2087-2096, 2001A.

Lo *et al., Genome Res. 11(3):* 448-457, 2001B

Murphy and Karpen, *Cell 82:* 599-609, 1995.

Page *et al., Hum. Mol. Genet. 4:* 289-294, 1995.

Paulson and Laemmli, *Cell 12:* 817-828, 1977.

Pinsky *et al., Dev. Cell. 3(1):* 4-6, 2002.

Pluta *et al., Science 270:* 1591-1594, 1995.

Pluta *et al., Trends Biochem. 15:* 181-185, 1990.

Saffery *et al., Hum. Mol. Genet. 9(2):* 175-185, 2000.

Saffery *et al., Proc. Natl. Acad. Sci. USA 98(10):* 5705-5710, 2001.

Saitoh *et al., Bioassays 17(9):* 2919-2926, 1995.

Schulze *et al., Mol. Gen. Genet. 264(6):* 782-789, 2001.

Steiner *et al., Mol. Cell. Biol. 13:* 4578-4587, 1993.

Su *et al., Proc. Natl. Acad. Sci. USA 99:* 4465-4470, 2002.

Sugata *et al., Hum. Mol. Genet. 9(19):* 2919-2926, 2000.

Sugata *et al., J. Biol. Chem. 274(39):* 27343-27346, 1999.

Sullivan and Schwartz, *Hum. Mol. Genet. 4:* 2189-2197, 1995.

Tomkiel *et al., J. Cell. Biol. 125:* 531-545, 1994.

Trowell *et al., Hum. Mol. Genet. 2:* 1639-1649, 1993.

Tyler-Smith *et al., Nature Genet 5:* 368-375, 1993.

Vagnarelli *et al., Chromsoma 110(6):* 393-401, 2001.

Van Hooser *et al., J. Cell. Sci. 114(19):* 3529-3542, 2001.

Wevrick and Willard, *Nucl. Acids. Res. 19:* 2295-2301, 1991.

Wevrick and Willard, *Proc. Natl. Acad. Sci. USA 86:* 9394-9398, 1989.

Wong *et al., Gene Ther. 9:* 724-726, 2002.